On the Performance and Effectiveness of Contact Tracing Apps in the Second Wave of COVID-19 in Italy

Leonardo Maccari, *Member, IEEE,*
Ca' Foscari University of Venice, Italy
Department of Environmental Sciences, Informatics and Statistics (DAIS)
leonardo.maccari@unive.it

Abstract—Contact tracing smartphone applications have been developed and used as a complement to manual contact tracing in the COVID-19 pandemic. The goal of these apps is to trace contacts between people and notify the mobile phone owners when one of their contacts tested positive. People who receive a notification should behave as *exposed* people, take a test and possibly isolate themselves till they receive the result. Unfortunately, identifying contacts based on distance is technically a daunting task: apps can be configured conservatively (a very small number of people is notified, limiting the effectiveness of the app) or they may be more tolerant and produce a high number of notifications but also of false positives. We review the data available from Immuni, the Italian app, which provides detailed figures on the notifications sent and the positive users, and we show that Immuni was configured to generate a very large amount of notifications. We estimate the testing resources that the health system would have needed if the app was downloaded by 100% of the adult population, and every notified person would require a test. In such conditions, Immuni would have generated a number of tests orders of magnitude higher than what available. We compare the performance of Immuni with the currently available literature on other apps and observe that contact tracing apps had a limited impact in the second wave of the COVID-19 pandemic. As contact tracing exposes citizens to privacy risks, we discuss some ways to reshape the goal of the apps to achieve a better trade-off between social benefit and risk.

I. INTRODUCTION

During the COVID-19 pandemic, manual contact tracing was one of the methods used to reduce the virus propagation. Manual contact tracing is based on interviews with people positive to the Sars-CoV2 aetiological agent and it is aimed at detecting the close contacts that might have been infected, testing them as soon as possible, and limiting the spread of the virus. In the attempt to speed up this process, which is time and resource-consuming, a plethora of contact tracing phone applications were developed, supported by works that suggested their theoretical usefulness [1]. These apps detect the proximity of nearby phones and can be used to send and receive notifications when a person tests positive. The design choice of these apps generated a discussion among privacy and security experts since this is, essentially, the first *state-sponsored mass tracing* ever introduced in democratic countries [2], [3]. As a result, an agreement was found (at the European level, but not only) on a *decentralized model* in which contact tracing is performed directly on mobile phones and only a minimal amount of data is shared with a centralized server [4]. This model minimizes the risk of leaks of private information, especially when compared with a fully centralized solution in which all data are stored in the server. Yet it does not remove risks, as contact tracing applications still collect sensitive information. Three key examples suffice i) who receives the contact notification should not know the identity of the person that tested positive, but the app stores the precise moment and duration of the contacts, and thus he/she may under certain conditions de-anonymize the identity of the infected person; ii) anyone in possession (or in control) of two mobile phones can reconstruct a precise history of meetings between the owners; iii) the centralized server that the app uses is a single point of failure, if compromised, an attacker could modify it and have access to a massive amount of identities that tested positive to COVID-19. These risks have been reviewed by the same DP-3T project that designed distributed contact tracing [5][1], and their recent technical feasibility has been verified on real applications [6], [7]. As tangible examples of the risk associated with digital contact tracing, it is worth mentioning the controversies sparkled in countries in which police was allowed to access data collected by contact-tracing applications [8].

Since risk can be mitigated but not removed, it is extremely important to show a social benefit that justifies the risk, assessing the effectiveness of digital contact tracing. Yet, there is a lack of both data and a methodology to perform such assessment. What is emerging from the latest literature is that the process of evaluating the impact of apps is split into at least two tasks. The first is to measure technical performance indicators that can tell if the app is working properly from a technological point of view: we always use *performance* when dealing with technical issues. The second is to understand how the output of the app produces a positive impact on public health, which is generally referred to as *effectiveness* in the medical literature (the positive impact of some intervention in a real-world situation). Both tasks are extremely challenging, as they require data that are hard (or even impossible) to access and their interpretation requires a cross-disciplinary approach.

---

[1]The reader may refer to the risks labeled as IR1, SR1, SR2, and the various risks associated with the server that receives the notifications from positive users.

While there is emerging literature in the epidemiological field that approaches this theme, this paper contributes to the discussion with a technical-oriented evaluation, based on the data coming from Immuni, the Italian contact tracing application, in the period of the peak of the second wave of the pandemic (September-November 2020). From the data, we can assess that Immuni was configured to maximize the number of identified contacts, at the expense of the precision of the detection. We then evaluate the effect of this design choice in terms of effectiveness showing that in an ideal situation (100% coverage at that time of the second wave) Immuni would have demanded an intolerably high amount of tests per day. The paper ends with a discussion on the possible reason for the negative performance of Immuni, and how these are naturally correlated with the choice of the technology, Bluetooth Low Energy (BLE), which is common to the majority of the apps. It also suggests potential ways to reshape the goal of the apps in order to take advantage of the large diffusion they reached among the population. To encourage replication (and possible falsification) of the results with data coming from other applications we share the source code and all the data[2].

II. STATE OF THE ART

The analysis of contact tracing apps is an extremely timely topic and the literature offers several works that classify the existing apps according to their technology, the architectural design, their privacy features [9], [10], [11], [12], [13]. Given the width and the depth of the coverage of this topic, it is not the goal of this paper to technically compare Immuni with the other applications; we will report only the details needed to support the analysis.

The effectiveness of contact tracing apps is a less explored theme. From a purely theoretical point of view, a contact tracing app that has enough accuracy, reaches enough diffusion, and delivers timely messages to people at risk has been shown to be useful in slowing down the pandemic [1]. In practice, the technology that is available and widespread makes it hard to reach a sufficient level of accuracy and diffusion in a real-world setting [14], [15], [16] and a discussion on apps utility at a policy level is undergoing [17], [18].

What makes assessing the effectiveness of contact tracing apps challenging is the lack of open data to base any analysis. Recent works adopt different strategies to route around this difficulty. One approach is to use surveys; some works showed that the users of the apps tend to react to the notification they receive [19] and that incentives may be effective to increase the applications uptake [20]. Yet, surveys have intrinsic limitations, such as the limited scale of the sample and the impossibility to represent the variability of the population adopting the app. Another approach is to try to reproduce on a small scale the contact tracing app [21] and feed a theoretical model with the numbers that are extracted from the experiments. However in this case the authors are not validating the app as they don't have a ground truth to verify

the accuracy of the contact tracing. Instead, they are using the app itself as their ground truth to estimate contacts, and thus, the spread of the virus. Another approach that is relevant in the discussion is taken by Wymant et. al with the estimation of the secondary attack rate (based on non-public data [22]), that leads to mildly optimistic conclusions. We will analyze that work in more detail in the next section.

III. DIGITAL AND MANUAL CONTACT TRACING

Let us recall that the overall goal of contact tracing is to reduce the reproductive number R_t, that is the average number of people that are infected by one infected person. If $R_t > 1$ then each positive person infects more than one person, so the contagion is expanding, else, the contagion is slowing down. Contact tracing tries to track *close contact events*, defined by several national health systems as a contact between two people without protections that lasts for at least x minutes at a distance of less than y meters. The values of x and y are based on scientific evidence that defines the contacts for which there is a high probability of infection. Albeit these numbers are debated [23] we assume there are thresholds beyond which the probability of infection decreases steeply, otherwise, the identification of close contacts based on distance would make little sense, and digital contact tracing with it.

Manual contact tracing requires an expert of the field that interviews the person that tested positive and is intended to reconstruct the history of *close contacts events* to identify the people that were exposed to the virus, the *close contacts*. Since even asymptomatic infected people are a vector of the contagion, close contacts are promptly quarantined to reduce the chances that they infect someone else. Quarantine generally ends when the person is tested negative, or if a test can not be done due to lack of testing capacity, after a certain number of days in which the quarantined person did not show any symptom. This has social and economic consequences and a limited throughput: we don't have an infinite capacity to perform tests and we should try to use it for those that have a higher risk of being infected.

A *detected close contact event* is instead the situation in which a mobile phone equipped with the app senses the presence of another mobile phone equipped with the app for more than x minutes at an estimated distance of less than y meters, performing *proximity detection*. Contact tracing apps were introduced with the assumption that a detected close contact is a good estimation of a close contact. The use of contact tracing apps has two potential positive effects, the first is to complement manual contact tracing increasing the number of close contacts that are detected. There are situations in which interviewed people do not remember or do not know about close contact events, like in public transportation. Moreover, manual contact tracing is time-consuming and the number of daily infected people may be higher than the number of people that can be interviewed. Automatic contact tracing based on smartphone apps should then increase the number of traced contacts. The second positive effect is to reduce the time needed to send notifications to close contacts, which is a key factor to fight the pandemic, because once a

---

[2]The code can be found at https://github.com/UniVe-NeDS-Lab/immuni-data-second-wave, together with the elaborated data. For the data sources see Appendix A

person tests positive, reducing the time to send the notification helps contacts to isolate earlier. Note however that while the first advantage is provided by the proximity detection function, the second advantage does not depend on it, but on a well-designed public health app that delivers timely notifications to close contacts (however they are identified). This is a key detail as the privacy issues of digital contact tracing reside in the risks associated with proximity detection, and they would mostly vanish if these apps could be transformed into generic notification apps. If instead digital contact tracing is meant to complement manual contact tracing, we need to focus on the performance of proximity detection.

IV. Technical Performance Indicators Vs App Effectiveness

Immuni uses the so-called Exposure Notification API developed by Apple and Google (EN, for short)[3]. EN is a software subsystem that enables any phone equipped with the Bluetooth Low Energy communication standard to perform proximity estimation with any other phone (more details in Sect. VIII-A). It estimates distance using the received power of the BLE packets (the so-called Received Signal Strength Indication, RSSI). EN is used by basically all the contact tracing platforms that want to reach mass diffusion, as the support from the Operating System is key to achieve high coverage, however, EN does not detect close contacts, it only estimates distance. Every app sets a minimum threshold received value t_{RSSI} so that if $RSSI > t_{RSSI}$ for a certain time, the contact is considered "close". Needless to say t_{RSSI} is a key performance parameter and strongly impacts effectiveness.

Apparently, the designers of apps took two separate directions and several adjustments on the way. Leith and Farrel [24] performed experiments with the Swiss, German and Italian apps that use different values for t_{RSSI} and showed that in a real-world scenario (public metro) the first two apps would detect zero contacts, and the third would detect anyone in the same coach with 50% probability, regardless of the distance. This is not surprising as the technical difficulty of measuring proximity using RSSI is well known to telecommunication engineers. An analysis of the scientific literature on proximity detection at the time when apps were proposed showed a lack of evidence to suggest that EN could be providing reasonable proximity estimations [16].

The developers of contact tracing apps then face a dilemma: given that the distance threshold to define a close contact changes with the environment (indoors/outdoors) and the scientific evidence correlated with the virus [23], and given that distance estimation is itself extremely noisy, should they adopt a high t_{RSSI} (that would generate a small number of contacts) or a low t_{RSSI} (that would generate a high number of false positives)? The answer to this question can not be given on a technical basis, but only considering the effectiveness of the app.

A. Estimating Digital Contact Tracing Effectiveness

One approach to estimate the effectiveness of digital contact tracing apps is to correlate at a local level the decrease of the positive cases with the penetration of the application. This is done by Wymant et al. [22] using the data coming from the British app, and showing that a higher penetration of the app correlates with a lower number of positive cases. The British app has several features besides contact tracing, so this observation needs to be complemented with other metrics in order to provide an estimation of the effectiveness of contact tracing. A second methodology used in the paper is to estimate the secondary attack rate (SAR) on the population of users notified, that is the fraction of them that later on tested positive. The authors estimated that the SAR computed for digital contact tracing is similar to the SAR computed on people that went through manual contact tracing, suggesting the two contact tracing have similar performance. This conclusion seems fairly optimistic for at least two reasons. First is they compare the estimated value of SAR with those the NHS (in its report of the period [25]) call *close contacts* (SAR=6.9) but they don't consider what instead the report refers to as *direct contacts*[4] (SAR=13.2). The two categories are clearly not disjoint, digital contact tracing may detect close *and* direct contacts, so a fairer comparison would be with the overall SAR of both categories, which is reported to be 12.7. This value is more than twice what is estimated for digital contact tracing. As a second reason, it is likely that some of the contacts that are detected by the app are the same that are traced with manual contact tracing, or that get automatically self-isolated (in the UK in the period, households of positive cases were automatically isolated). If the overlapping is high, then the positive impact of the app is strongly reduced, as the app is providing contacts that are already known, and it is limited to the speed-up in the notification process (for the ones that are not immediately reachable by the positive person, e.g. households). This is only briefly analyzed in the paper, and the necessary data to verify it are not available, but it is instead a key point to evaluate the effectiveness of digital contact tracing applications, as SAR only is not a good indicator.

According to the recent guidelines by the World Health Organization on the evaluation of contact tracing apps [26] effectiveness should be evaluated as the *"Proportion of diagnosed cases previously notified only through the app (but not through conventional contact tracing) among all diagnose"*. Unfortunately, this is extremely hard to automatically assess, because it implies that when a person has a positive test, he/she declares that the reason for taking the test was an app notification and not any other. The only way to reconstruct this chain of events is with surveys, which are of limited size and time duration.

We use a different approach based on publicly available data that shows how the performance indicator we use to evaluate Immuni reflects in an estimation of effectiveness. The goal of this analysis is to put under the spotlight that a bad design

---

[3]Privacy-Preserving Contact Tracing, see https://www.apple.com/covid19/contacttracing/. Visited 11/09/2021.

[4]Defined as *"face to face contact (for example a conversation within 1 metre); skin to skin contact (including sexual contact); coughed on, sneezed on or spat on."*.

choice could produce effectiveness metrics that make the app completely unusable and unsustainable, as was the case with Immuni during the second wave.

V. EVALUATING IMMUNI

Since fall 2020 the developers of Immuni started to publish open data that can be used to estimate the precision and the scalability of the app[5]. To our knowledge, Immuni is the only app that publishes the number of notifications sent by every app and the number of infected people that sent a notification.[6]

We focus on the period of the emergence of the second wave, which we define to be between Aug. 13th (the week in which R_t became larger than 1) and Nov. 3rd (the beginning of the lockdown measures). Fig. 1 reports the number of daily new infected people in Italy in fall 2020, together with the estimated value of R_t obtained from official sources. In this time interval, manual contact tracing could not cope with the high number of infected people per day (a peak of 40,902 on November 13th) as the resources needed to interview such a huge number of people and notify the close contacts were unavailable. We are interested in the performance of Immuni in this period, because the social benefit of digital contact tracing could have been maximal but also because after Nov. 3rd two events changed the scenario. First, national authorities enforced lockdown measures and Immuni started to operate in an abnormal situation in which only a very small number of contacts were legally allowed. Second, the Immuni developers confirmed (without providing a rationale) that on Oct. 28th t_{RSSI} was raised (see Appendix C). Data show that after that date, the number of identified contacts sharply decreases and the performance is not comparable with the previous period. Our observations and conclusions then refer to the second wave period, highlighted in all the figures, but when relevant we also report the metrics for some weeks after Nov. 3rd. Relevant symbols used in the paper are summarized in Tab. I.

A. A Technical Performance Indicator: notifications per positive case

One of the key technical performance indicators mentioned by the WHO [26] for contact tracing apps is the ratio between notifications and positive registered cases averaged over a week time. If we call $i(t)$ the number of people that in a certain day had a positive test, $i'(t)$ the numbers of them that were running the app and notified to the app that they tested positive, and $S'(t)$ the total number of notifications sent in a day by the app, $\alpha(t)$ is defined as:

$$\alpha(t) = \frac{S'(t)}{i'(t)} \qquad (1)$$

provides a rough estimation of how the designers of the app configured the threshold of the proximity estimation function.

[5]For readability, all the URLs needed to access the datasets and the source code for the analysis are reported in Appendix A.

[6]The German Corona app also publishes the number of notifications, but only for a subset of users that voluntarily agreed to share their data. This decouples the numbers of positive cases from the numbers of notifications and does not allow us to repeat the same analysis we do for Immuni.
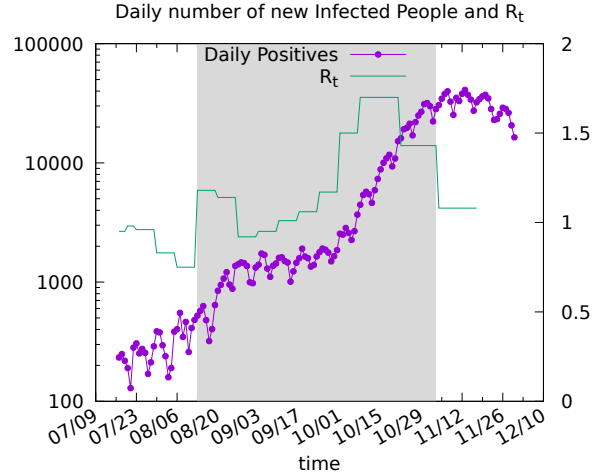

Fig. 1. Number of newly infected people per day (left axis) and R_t (right axis).

We will compute the weekly average of $\alpha(t)$ and compare it with the same value reported by the UK application in roughly the same period of analysis.

B. Consequences on Effectiveness

Let us consider a situation in which the app reached 100% coverage among the population that is allowed to use the app and that the app never fails to identify a contact. We refer to this as the *ideal situation*, in which the app reaches total diffusion and identifies all the contacts.

Let us call $n(t)$ the average number of detectable close contacts that an infected person had in the period that is considered for contact tracing. The value of $n(t)$ depends on the time window that is adopted for contact tracing (normally 2-3 days before the emergence of the symptoms), on the behavior of people, and thus on the restrictions they are subject to. Since restrictions change with time, $n(t)$ has an explicit dependency on time t, for which we use a day granularity. $n(t)$ is unknown and we will use the data from Immuni to estimate it. Let us now consider the following metric:

$$\gamma(t) = \frac{R_t}{n(t)} \qquad (2)$$

From a purely technical point of view $\gamma(t)$ is a precision metric. If we consider that R_t is the average number of contacts that we are looking for (the number of true positives), and that $n(t)$ is the average number of contacts that the app will provide (true positives + false positives), then we could treat $\gamma(t)$ as a Positive Predicted Value and expect it to be as close as possible to one. If $\gamma(t)$ is smaller than one, the app sends notifications to people that are not at risk, and this is a technical failure for the app. From a usability point of view, it is clear that precision plays a key role because being self-isolated for a false positive notification influences the trust people have in the app itself and encourages/discourages people to use the app, as the recent discussion on the so-called "*pingdemics*" showed in the UK [27].

However, from an epidemiological point of view this approach is not valid, as contact tracing is not meant to identify only the positive contacts, but a group of people that is large enough to contain all the positive contacts with high probability. In this sense, effectiveness is impossible to evaluate, as there is no ground truth to compare with (we don't know the real number of people to be considered at risk). Yet, as a minimum benchmark, digital contact tracing should perform better than picking people at random. We then compare $\gamma(t)$ with the estimated infection incidence in Italy in the same period, that is the number of infected cases per 100.000 people, estimated from $i(t)$. Viceti et al. [28] take the number of reported cases in the second wave and scale it by a factor coming from the Seroprevalence Investigation on SARS-COV-2 realized by the Italian National Health Institute in July/August 2020. The investigation tested a representative portion of the Italian population and reported that the number of real cases was 6 times the number of identified ones. Viceti et al. thus report that the average incidence for the whole period was 679/100,000. We use this number as a baseline comparison for $\gamma(t)$ and adopt a similar approach to have a time-varying trend: for every Italian region, we take the number of cases identified in a 14-days window and rescale it by a factor of 6. We compare $\gamma(t)$ with the range of values provided by the region with the minimum and maximum incidence. The 14 days interval was chosen as Immuni stores data for two weeks, so it can detect contacts as far as 14 days after the contact itself.

Finally, to compute $\gamma(t)$ we need an estimation of R_t. In the period under analysis (July-November 2020) the official sources estimated R_t to be 1.7 at its peak [29] (see Fig. 1), while recent studies estimated the value of R_t in the second wave to be 2.8 [30], with a pretty large 95% confidence interval (1.5-4.2). We use both values in order to compute $\gamma(t)$ on a range of realistic values.

C. Impact on testing capacity

We call $\overline{S}(t)$ the overall number of contacts that an app will detect at day t in the whole country assuming the ideal situation. Since every person that receives a notification from the contact tracing app is at risk, he/she should be tested, then $\overline{S}(t)$ provides the number of tests that are needed every day. We compare $\overline{S}(t)$ with the real capacity of making tests from the Italian health system (at the time of the second wave) to provide an estimation of how effective digital contact tracing is, given the resources we can employ against COVID-19. More formally, given that $T(t)$ is the reported number of tested people per day, our second performance metric is given by:

$$\rho(t) = \frac{\overline{S}(t)}{T(t)} \tag{3}$$

The $\rho(t)$ metric represents the ratio between the tests that would have been necessary and the tests that were daily done by the national health system. If $\rho(t)$ is larger than one, the contact tracing app requires more testing capability than what is available.

D. Discussion on the metrics Assumptions

Both R_t and $n(t)$ in $\gamma(t)$ are average absolute numbers and so they must be compared in the same conditions. R_t is the number of people infected by one person considering all the national population as the group of the susceptible people, and thus, $n(t)$ represents the contacts we expect that could be detectable in the ideal situation. If we drop the 100% assumption then $n(t)$ decreases because fewer contacts would be detectable, but then we should rescale also R_t of the same fraction, so $\alpha(t)$ would be unchanged (or else, we would get to the absurd that decreasing the penetration, the precision increases). For simplicity, we keep the 100% penetration assumption and we don't consider the false negatives.

Note also that referring to the effectiveness definition provided in Sect. IV-A the definition of $\gamma(t)$ is "optimistic" for Immuni, because a majority of the contacts that test positive falls in very predictable groups (primarily households [31], [16]) that can be isolated with manual contact tracing or simple compulsory isolation for households of infected people. So in general, the number of people that the app should help to identify is lower than R_t, because other tracking measures are in place.

Another important assumption is that every person that receives a notification will take a test, which is not strictly mandatory. In general, people that receive a notification should behave as people at risk, self-isolate, and follow the same procedure of contacts that were manually traced. A person may decide to take a test or wait for the maximum length of the self-isolation without making a test, and in case there are no symptoms he/she can exit isolation anyway. Under the point of view of social impact, a test that is officially recognized as valid (PCR or rapid antigen tests) is the method that limits the damage the most and comparing the number of required tests with $T(t)$ is a very intuitive way of measuring effectiveness. Alternatively, we should consider several possible outcomes of the reception of a notification, and use a model that takes into consideration the positive and negative impacts or a mix of testing, self-isolation, quarantine, etc. Some works go in this direction [32] but they use data that are not available nationwide (like a reasonable estimation of the real contact graph) and in general, this kind of analysis is out of the scope of this paper. What we want to stress is the total disproportion between the metric values and baseline comparison, so we rely on a very intuitive metric.

Finally, it must be noted that at the time of writing the testing capacity increased two folds compared with the second wave. When commenting on the results we note that even if we had that testing capacity at the time, $\rho(t)$ was so high that our conclusions would still hold.

VI. ESTIMATING THE IMPACT OF IMMUNI

We call $C(t)$ the probability that the app runs on someone's smartphone. $C(t)$ depends on several parameters, as a first approximation we can say that:

$$C(t) = M \times A(t) \times W \tag{4}$$

where M is the fraction of the population owning a smartphone, $A(t)$ is the fraction of the population that installed

Evaluation Metrics

$\alpha(t)$	Warnings per positive case: $\alpha(t) = \frac{S'(t)}{i'(t)}$
$\gamma(t)$	R_t over estimated contacts: $\gamma(t) = \frac{R_t}{n(t)}$
$\rho(t)$	Number of tests over testing capacity: $\rho(t) = \frac{\overline{S}(t)}{T(t)}$

Data from official sources or literature

R_t	Reproductive number of the virus ([29], [30])
$T(t)$	Reported number of tested people per day (IHM)
M	Fraction of the population owning a smartphone (IHM)
$i(t)$	Number of people that tested positive for the first time in day t (IHM)
$D(t)$	Number of downloads (Immuni data-set)
W	Ratio between downloaded apps and running apps (Swiss-Covid App)
P	Average probability of correct detection (set to 1 in the ideal case)
$i'(t)$	Number of infected users running the app (Immuni data-set)
$S'(t)$	Number of notifications at a certain time t (Immuni data-set)

Derived metrics

$n(t)$	Average number of detectable contacts (estimated)
$\overline{S}(t)$	Expected number of contacts detected nation-wide assuming 100% penetration
$C(t)$	Probability that the app runs on someone's smartphone: $C(t) = M \times A(t) \times W$
$A(t)$	Fraction of the population that installed the application once: $A(t) = \frac{D(t)}{R}$
R	Number of people allowed to download the application
$S(t)$	Overall number of contacts detected at t
$N(t)$	Overall number of detectable close contacts nation-wide for day t

TABLE I
REFERENCE FOR THE SYMBOLS USED IN THE PAPER. IHM STANDS FOR ITALIAN HEALTH MINISTRY, SEE APPENDIX FOR DETAILS AND LINKS.

the application once (computed as the number of downloads $D(t)$ divided by the number R of people that are allowed to download the application due to age restrictions), W is the ratio between downloaded apps and running apps, which takes into account the fact that not every download corresponds to a running app. M and W depend on demographics, market and technical limitations of phones and operating systems, i.e., factors that do not change in the observed period. $A(t)$ instead has a time dependency as the number of people that downloaded the app grows with time.

Being $i(t)$ the number of people that tested positive for the first time in day t, we define

$$N(t) = n(t) \times i(t) \qquad (5)$$

that is the daily number of detectable close contacts nation-wide. If the app has perfect accuracy, under the assumption of 100% penetration, it will detect all these contacts. We call P the average probability that an app running in one phone can detect the presence of another app in a phone in proximity. P depends on the Bluetooth technology and on the t_{RSSI} parameter. A contact is traced with probability P when both people involved are running the app so the average probability of detecting a close contact is $C^2(t) \times P$. Finally, we call $S(t)$ the overall number of contacts that will be detected daily at time t in a certain country, for which it holds:

$$S(t) = N(t) \times C^2(t) \times P = n(t) \times C^2(t) \times P \times i(t) \qquad (6)$$

meaning that the total number of daily detected contacts will be given by the average number of detectable contact per person, multiplied by the probability of tracing each of them, and scaled on the number of newly infected people in the country at time t.

Our first goal is to estimate $n(t)$ using the available open data from Immuni, which provides two important data entries: the total number of infected users running the app[7], and the total number of notifications at a certain time t (we call them $i'(t)$ and $S'(t)$, respectively). Note however that when we collect data for a specific app, we get to know the number of infected people among those *running the app* so the probability that one of their contacts is detected becomes $C(t) \times P$, that is, the probability that the other person is running the app multiplied by the accuracy. Thus, similarly to Eq. (6), we can write:

$$S'(t) = n(t) \times C(t) \times P \times i'(t) \qquad (7)$$

From which we derive $n(t)$

$$n(t) = \frac{S'(t)}{C(t)i'(t)P} \qquad (8)$$

To obtain $n(t)$ from Eq. (8) we use $S'(t)$ and $i'(t)$ from the open data and we need a reasonable estimation of $C(t)$ which we will provide in the next section. Once we have $n(t)$ we can already compute the $\gamma(t)$, which is the first metric we are interested in.

The second metric $\overline{S}(t)$, is the value of $S(t)$ in the specific case of 100% coverage, that can be obtained using Eq. (6) in which we plug the value of $n(t)$ and we set $A(t) = W = 1$ to enforce the 100% coverage condition.

A. Estimating C(t)

In order to estimate $C(t)$, we need values for all the elements of Eq. (4). To estimate W we use the data provided by the SwissCovid App website, the Swiss application for contact tracing that uses an in-app notification to estimate the number of running apps. Such data provide an estimation of the effective number of applications compared to the downloads, a difference that might be due to technical reasons (the app does not run on a certain version of hardware or OS) or to users that simply uninstall the app. SwissCovid uses the EN subsystem as well as Immuni, so we can assume that the technical reasons preventing its usage are common to all the apps that use EN. We then use the SwissCovid App data as a reference also for the Italian case. At the time of writing the app has been downloaded 2,768,968 times and it is actively used by 1,860,000 users so we set $W = 1,860,000 \times 2,768,968^{-1} = 0.67$.

We then use data from the Pew survey on mobile phone penetration to estimate M. Pew reports that 91% of the Italians older than 18 own a smartphone. Immuni can be downloaded only by people aged at least 14 so the sample does not

---

[7]Note that notification on the app is not compulsory, so what Immuni reports is the total number of positive people owning the app that notified it to the app itself. For simplicity, we will use the "infected users" term.
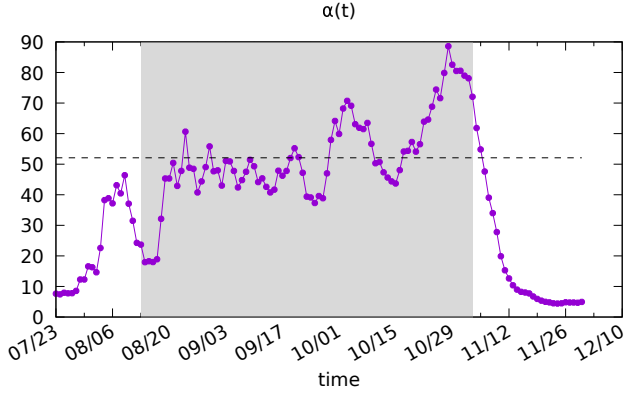
Fig. 2. The value of $\alpha(t)$ (notifications per positive cases), weekly average (solid, dotted line), and the overall average for the second wave period (dashed line).

Fig. 3. $\gamma(t)$ metric (R_t over estimated contacts) for Immuni in the reference period, computed using R_t =2.8.

perfectly match the population or Immuni users, but in the absence of other data, we set $M = 0.91$. The Italian population older than 14 is reported to be 52.997.219 by the national statistical institute, so we set $A(t) = D(t) \times (52,997.219)^{-1}$ where $D(t)$ is the number of downloads reported by Immuni.

The Italian national health system provides the number of daily new infected people (summarized by Arisi and Mantuano [33]). Among the whole number of infected Italians, the ones of age between 0 and 9 and between 10 and 19 correspond to 8.44% and 9.58% respectively. In absence of a finer-grained statistic, we remove $8.44 + \frac{4}{10}9.58 = 12.27\%$ from the total and we set $i(t)$ to be $87,73\%$ of the total number of daily infected people.

Finally, since we consider the ideal situation and know that Immuni set t_{RSSI} at the minimum level, we set $P = 1$, assuming that the probability of not receiving any signal from a contact is low.

Immuni provides the number $i'(t)$ of positive users and the number of notifications that were sent to people that were in contact with some positive person $S'(t)$. Since $T(t)$ has a weekly cycle (during the weekend fewer tests are performed) we extract from the Immuni open data the moving average on a window of 7 days of $S'(t)$, $i'(t)$ and $A(t)$, we use the latter to compute $C(t)$ with the given M and W, and we plug these values in Eq. (8), from which we obtain $n(t)$. Once we have an estimation of $n(t)$ we use it with $i(t)$ in Eq. (6) to determine $\overline{S}(t)$. We now have all the necessary ingredients to estimate $\gamma(t)$ and $\rho(t)$.

VII. RESULTS

During the rise of the second wave a total of 2451 Immuni users notified to be positive (weekly moving average 189) and this produced 162615 notifications (weekly moving average 12668). The median number of notifications per positive case was 52.1, and the average value of $\alpha(t)$ is reported in Fig. 2. The average value of $\alpha(t)$ reported by Wymant et al. for the
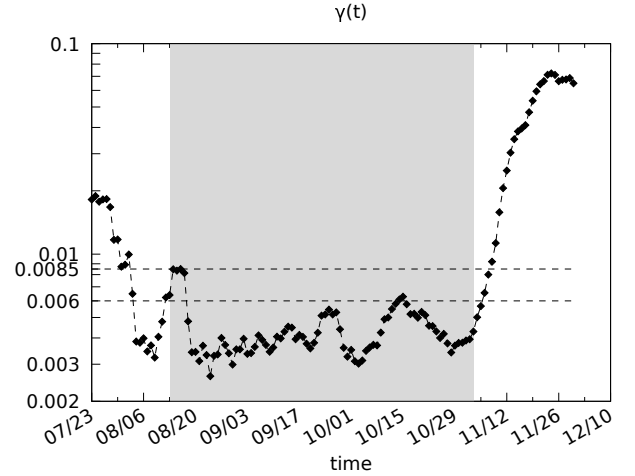
raise of the second wave in the UK is 4.2, which is an order of magnitude lower. This confirms that the developers of Immuni have used, till Nov. 2020 a very low t_{RSSI}, that produced an extremely high number of notifications per positive user.

Fig. 3 shows that during the rise of the second wave, when Immuni could have produced the highest benefit, its $\gamma(t)$ was constantly lower than 0.0085 (0.85%) and in 84% of the days lower than 0.006. In the highlighted period the average $\gamma(t)$ was 0.0043 which means that on average among all the people that could receive a notification, 0.43% would have been tested positive.

We compare $\gamma(t)$ with the estimated incidence of the virus due to the new infections in the second wave only. Fig. 4 reports a zoom of the data in the referred period and reports two trends. The first one (green color) is the envelope between $\gamma(t)$ when estimated with R_t =1.7 (lower bound) and R_t =2.8 (upper bound). The dashed curve is the average incidence of new cases in a 2 week period estimated using the official sources, the blue area is bounded by the highest and lowest incidence reported by the Italian regions. The figure does not provide any strong evidence suggesting that digital contact tracing provides a higher number of contacts than random sampling the population.

Finally, the value of $\rho(t)$ in Fig. 5 shows that at the peak of the contagion the number of tests we would have needed in ideal conditions was two orders of magnitude higher than what we could afford, and at its minimum, it still required at least 3.5 times the available testing capacity. The average $\rho(t)$ in the period of the second wave was 26.6, including the weeks after the lockdown the average lowered to 22 and always stayed larger than 5. If we compare it with the national population, we see in Fig. 6 that the required number of tests would have been sufficient to test all the Italian adult population every three days. If we had that testing capacity, we would have not needed any contact tracing at all, because all the people would have been constantly monitored.

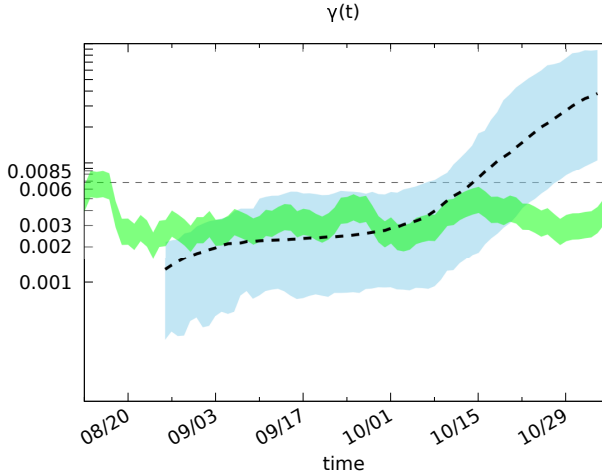It is interesting to note that during the second wave the

Fig. 4. The comparison between $\gamma(t)$ computed with two values of R_t (green area) and the incidence estimated using official data (blue area). The blue area is limited by the maximum and minimum incidence estimated on Italian regions and the dashed black line is the average. The dashed gray line is the average incidence computed on the whole second wave period as reported in [28].
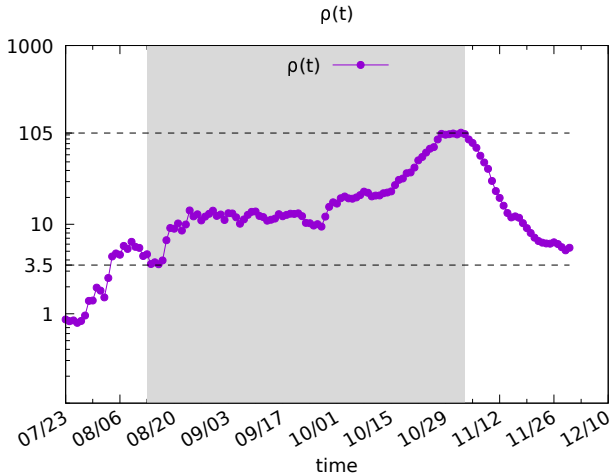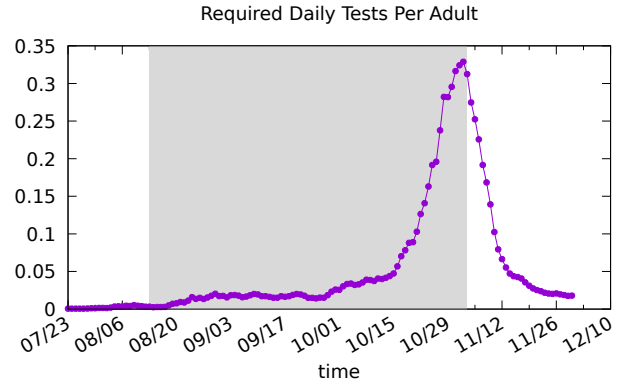

Fig. 6. Estimated average number of tests per adult with 100% coverage.

reference to the EN platform. We then try to imagine how this technology could be changed to play a different role, exploiting the momentum it gained, with millions of people that downloaded it.

A. Estimating Distance with BLE

The EN technology uses the Bluetooth Low Energy standard to estimate the distance between two phones, here we report just a brief overview of its functions, a more accurate review can be found in the references [10].

When using BLE the Bluetooth radio of the phone remains turned off most of the time, then periodically it switches on and sends broadcast *beacons*, i.e. short packets that signal the presence of the phone to the devices that are in the communication range. This happens, according to the EN specifications, at most 4 times per second. With a different period (order of minutes), the radio switches on and listens for incoming beacons to detect nearby devices. Phones do not engage in any *handshake* of packets, they just keep broadcasting a beacon whose content does not depend on the receiver. Knowing the RSSI and the transmission power, which is reported by the sender, the receiver app computes the attenuation of the electromagnetic wave and uses well-known path loss calculation models to estimate the distance [24].

The main issue with this technique is that RSSI is extremely noisy in real-world environments. There are many physical reasons why the received power could be lower or even larger than expected. These reasons depend on a large number of factors that are impossible to control, which deal with the precision of the radio measure, the surrounding environment which shades the signal and/or produces reflections that add up to the signal in line-of-sight, and of course, on the position of the phone (in a pocket, in a purse, etc.). To improve the estimation Google has performed tests on a large number of devices in a protected environment. It is interesting to note that from the Google documentation it emerges that even changing the orientation of just one of the devices may lead to a 10 dB difference in the received signal strength, which corresponds

Fig. 5. $\rho(t)$ metric for Immuni in the reference period.

testing capacity $T(t)$ passed from an average of about 39.000 per day in the first week, to about 157.000 in the days of the peak (the week between October and November), yet Fig. 6 shows that $\overline{S}(t)$ grew much faster than that. At the time of writing the testing capacity increased twofold due to the use of rapid antigen tests[8]. Even if we had twice the available testing capacity during the second wave, at its peak, we would still need more than 50 times the number of available tests.

VIII. DISCUSSION AND WAY AHEAD

In this section, we analyze the reasons behind the design choices of Immuni that lead to its poor performances, with

[8]According to official data, up to Sept. 9th, 2021 in the week with the highest number, about 329.000 tests per day were made.

to an error of meters[9]. In real life the conditions vary abruptly: the received signal may change dramatically from one room to another in the same building, which makes it impossible to model an *average* behavior applicable to all devices and all places. We already mentioned experiments that have shown that the accuracy of distance estimation is extremely poor in the Italian, Swiss, and German contact tracing applications [24]. The authors noticed that in the setting in which they performed the experiment there was no appreciable correlation between distance and the measured RSSI. They also confirmed that Immuni was configured to detect the highest number of contacts, which explains the high number of false positives we estimated.

Other techniques are known to be more accurate to estimate distance, for instance, measuring the time-of-flight (the time it takes for a packet to reach the receiver) has been shown to be helpful to reduce the error of distance estimation with BLE [34], but this technique requires a closed-loop handshake between the two devices. This *connected* model is possible on BLE [10] but has the main disadvantage that the number of packets sent is not predictable. In areas with a high density of devices, the phones could create a *storm* of packets if for some reason they decide to perform the handshake at the same time (e.g. when a new device is switched on). This has nefarious effects on precision (more packets imply more noise and less precision) and most of all, on energy consumption. The connected model sends and receives a number of packets that depends on the number of phones nearby, which makes energy consumption unpredictable. Since smartphones are energy-hungry, a background app must have a predictable and small energy footprint, or else its adoption would be hindered. The simplicity of the broadcast approach guarantees this, while the connected approach does not.

A review of the literature on contact tracing [16] with BLE shows that to have a decent estimation of the distance we need data from both terminals, and we need to train a system using some ground truth, and as we said, the ground truth changes from place to place. When analyzing the literature there were no hints that such a system would be working properly at a mass scale. The experiments that provided positive results were realized in settings that are not even closely comparable (i.e. more favorable) to the case of a mass contact tracing on user devices.

On the other extreme of the spectrum of Immuni, there are contact tracing apps that in the same period were configured with high t_{RSSI} and produced a lower number of notifications per contact. Besides the already mentioned UK application, recent works seem to estimate an increase of about 5% of the detected contacts for the Swiss [35] and Washington State app [36], while a Dutch report mentions a decrease of the R_t in the order of 0.3% due to the app [37]. These numbers must be confirmed using a standard methodology, but for the time being, we can say that digital contact tracing is not a game-changer in the fight against the pandemic. After more than one year of its adoption, it is time to compare its cost-effectiveness

with less privacy-intrusive measures.

Nevertheless, national states made a large effort to convince citizens to use them, and in fact, they are installed in tens of millions of phones, especially by those categories of people that are more at risk [20]. In the rest of this section we foresee two ways, and the related research challenges, that can be used to exploit this motivated user base to continue fighting the current, and possibly future pandemics.

B. Improve Proximity Detection

There are two ways to improve proximity detection, the first one is to use time-of-flight as mentioned, but as we said it would be hard to implement maintaining the benefits of the simple broadcast solution. The second option is to use a fully centralized system, as in all the experiments in the literature that achieved a satisfactory precision of encounters detection. A centralized server can use the data from both the endpoints of the communication and better compensate for the errors of one of the two, moreover, it could use the history of data to tune the sensitivity of the distance estimation, and to apply unsupervised learning techniques. If coupled with the geographic position of the phones, the algorithm parameters could change as a function of the specific environment where the encounter takes place, for instance differentiating between outdoor and indoor space. Data could be enriched using fixed anchors, i.e. nodes that are placed in a strategic place and can be used as a "ground truth". All these changes in the way the app and its back-end were designed would make the detection much more accurate than how it is now, and achieve results that are more in line with the ones that can be found in the literature in lab situations or controlled environments.

Unfortunately, such a system would lose all the privacy features that the current one has, and that were introduced after a public debate on the theme. Indeed it would be an extremely invasive instrument that centrally logs all the interactions between people, their position, and thus, in general, their habits. This was rightly considered unacceptable at the beginning of the pandemic. The system would need to be transformed into a privacy-aware system exploiting state-of-the-art cryptography concepts like k-anonymity [15], zero-knowledge proof [38], or homomorphic encryption [39], [40], [41]. While this is a very stimulating research challenge, it is far from being solved in the short term.

C. Risk Profile, Monitoring, and Nudging

Fig. 7 shows the trends of downloads of Immuni in the period of the second wave and shows a clear correlation between the number of downloads and the number of positives, meaning that, especially in October when the number of infections skyrocketed, people reacted by downloading the app. Not only, the initial analysis of the reactions of the SwissCovid App shows that people tend to react to the notifications that the app delivers [19].

These two observations suggest that such apps play both an emotional role, as a way that people use to increase their sense of safety and contribute to a possible solution, and are also considered a trusted source of information. Moreover, apps

[9]See https://developers.google.com/android/exposure-notifications/ble-attenuation-procedure#device-orientation. Visited 11/09/2021.
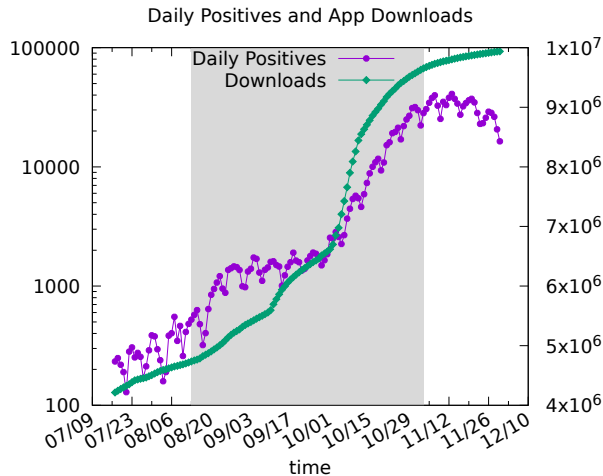
Fig. 7. Number of Immuni downloads and number of newly infected people per day, second wave of the pandemic.

are gaining features that make them appealing beyond contact tracing, for instance, the UK app is used to deliver results of tests, and Immuni is used also to obtain the Italian Green Pass (a QR code that certifies the vaccination status and allows to avoid a growing number of restrictions on public activities). Finally, the apps could help break the national barrier if their databases could be federated.

Unfortunately, our results indicate that a wrong design choice could produce an intolerable number of false-positive notifications and mine the trust in the apps. Every time a person decides to self-isolate and take a test because he/she received an app notification which later on reveals to be a false positive, the trust in the app will lower, till the moment in which notifications will be just ignored or the app uninstalled. In a period in which the uncertainty about the future leads to strong negative emotional consequences on society [42] and the spread of fake news about covid makes it hard to deliver a correct message to the general population [43], the fact that millions of people decided to keep a trusted source of information in their pocket is a chance that should not be wasted.

One way to positively exploit this situation is to change the goal of the app from a contact tracing app to an app that profiles the behavior of the user and provides suggestions and *nudges* on measures he/she can take to reduce his/her risk of being infected and infect others. For this application we need research results from a variety of fields, we limit to the analysis of two key components in the ICT domain. The first component is a profiling strategy that takes in input a series of factors regarding the user, including estimated contacts, general health conditions, age, geographical position, and habits. Fenton et al. for instance, propose to train a Bayesian network that can compute a risk profile for the user, based on several parameters [44], [45]. The app then elaborates feedback for the user on his/her health state, with indications on how to change one's behavior in order to reduce the risk itself. The second component is a way to design

an app that is effective in delivering feedback to the user, convincing him/her to take some action (limit interactions, consult a doctor, or take a test). An example of research in this field comes from Munzert et al. that introduce a design based on the widely used concept of *nudges*, interactions with the user aimed at providing suggestions on his/her behavior [20]. In the long run, the app would become a monitoring device that exploits the knowledge of private information (that remains local) to estimate the conditions of the owner. The output of the application would not be binary anymore (whether or not a close contact happened) but a risk profile, with suggested actions to reduce the risk. The app would mix local data with situation awareness, and thus behave differently depending on the state of the pandemic and on the position of the user. Finally, it would be possible to update the functions of the app to new needs when the current pandemic is over, for instance, to monitor seasonal flu and reduce the number of deaths it produces.

An app that produces risk profiles has the advantage that it does not need high precision in estimating distance and not even high penetration and finally, compared to centralized contact tracing described in Sect. VIII-B, this solution is not privacy-invasive, as most elaboration can be local and the model can be trained with aggregated and properly anonymized data provided by the users that test positive.

IX. CONCLUSIONS

In the early stage of the pandemic, digital contact tracing was introduced as an added instrument to fight the contagion. The underlying assumption was that the technology was precise enough to perform contact tracing. Immuni uses the EN proximity detection platform provided by Google and Apple, which is in use in the large majority of the contact tracing apps, and that has been tuned using the data from millions of users. EN represents the best of what is possible to achieve with current technology, under the constraints of a privacy-respecting approach. Yet, the final decision on the contact detection depends on configurations that change from one app to another.

Immuni during the second wave has shown to be extremely imprecise with an intolerable amount of false positives. Further data coming from other apps that used a more conservative configuration do not provide solid evidence that digital contact tracing had a key effect. In light of this, we should reconsider the design of contact tracing apps because more than one year after their mass adoption there are no indicators that suggest that the privacy risk is justified by a high social benefit. We could instead exploit the popularity of the apps and the trust they gained, refocusing their goal to achieve a higher social impact.

X. THANKS

REFERENCES

[1] L. Ferretti, C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner, M. Parker, D. Bonsall, and C. Fraser, "Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing," *Science*, vol. 368, 2020.

[2] K. Bock, C. R. Kühne, R. Mühlhoff, M. R. Ost, J. Pohle, and R. Rehak, "Data protection impact assessment for the corona app," 2020, visited 16/09/2021. [Online]. Available: https://www.fiff.de/dsfa-corona

[3] H. Cho, D. Ippolito, and Y. W. Yu, "Contact tracing mobile apps for COVID-19: Privacy considerations and related trade-offs," 2020, preprint. [Online]. Available: https://arxiv.org/abs/2003.11511

[4] C. Troncoso, M. Payer, J.-P. Hubaux, M. Salathé, J. Larus, E. Bugnion, W. Lueks, T. Stadler, A. Pyrgelis, D. Antonioli *et al.*, "Decentralized privacy-preserving proximity tracing," 2020, visited 16/09/2021. [Online]. Available: https://github.com/DP-3T/documents/blob/master/DP3T%20White%20Paper.pdf

[5] The DP-3T Project. (2021) Privacy and security risk evaluation of digital proximity tracing systems. Visited 16/09/2021. [Online]. Available: https://github.com/DP-3T/documents/tree/master/Security%20analysis

[6] D. J. Leith and S. Farrell, "Contact Tracing App Privacy: What Data Is Shared By Europe's GAEN Contact Tracing Apps," in *IEEE Conference on Computer Communications (INFOCOM)*, May 2021.

[7] L. Baumgärtner, A. Dmitrienko, B. Freisleben, A. Gruler, J. Höchst, J. Kühlberg, M. Mezini, R. Mitev, M. Miettinen, A. Muhamedagic, T. D. Nguyen, A. Penning, D. Pustelnik, F. Roos, A.-R. Sadeghi, M. Schwarz, and C. Uhl, "Mind the gap: Security amp; privacy risks of contact tracing apps," in *IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, Dec. 2020.

[8] T. Jamie. Governments tap covid data for other uses, risking backlash. Visited 16/09/2021. [Online]. Available: https://www.bloomberg.com/news/articles/2021-01-31/governments-exploit-covid-data-for-other-uses-risking-backlash

[9] N. Ahmed, R. A. Michelin, W. Xue, S. Ruj, R. Malaney, S. S. Kanhere, A. Seneviratne, W. Hu, H. Janicke, and S. K. Jha, "A survey of covid-19 contact tracing apps," *IEEE Access*, vol. 8, 2020.

[10] M. Cunche, A. Boutet, C. Castelluccia, C. Lauradoux, D. Le Métayer, and V. Roca, "On using Bluetooth-Low-Energy for contact tracing," Inria Grenoble Rhône-Alpes ; INSA de Lyon, Tech. Rep., Jun. 2020.

[11] V. Kouliaridis, G. Kambourakis, E. Chatzoglou, D. Geneiatakis, and H. Wang, "Dissecting contact tracing apps in the Android platform," *PLOS ONE*, vol. 16, no. 5, May 2021.

[12] L. Reichert, S. Brack, and B. Scheuermann, "A Survey of Automatic Contact Tracing Approaches Using Bluetooth Low Energy," *ACM Transactions on Computing for Healthcare*, vol. 2, no. 2, Mar. 2021.

[13] A. S. Krishnan, Y. Yang, and P. Schaumont, "Risk and Architecture Factors in Digital Exposure Notification," in *International Conference on Embedded Computer Systems*, July 2020.

[14] I. Braithwaite, T. Callender, M. Bullock, and R. W. Aldridge, "Automated and partly automated contact tracing: a systematic review to inform the control of COVID-19," *The Lancet Digital Health*, vol. 2, no. 11, Nov. 2020.

[15] J. Gonzalez-Cabañas, A. Cuevas, R. Cuevas, and M. Maier, "Digital Contact Tracing: Large-Scale Geolocation Data as an Alternative to Bluetooth-Based Apps Failure," *Electronics*, vol. 10, no. 9, Jan 2021.

[16] L. Maccari and V. Cagno, "Do we need a contact tracing app?" *Computer Communications*, vol. 166, Jan. 2021.

[17] A. Blasimme and E. Vayena, "What's next for COVID-19 apps? Governance and oversight," *Science*, vol. 370, no. 6518, Nov. 2020.

[18] L. C. Ivers and D. J. Weitzner, "Can digital contact tracing make up for lost time?" *The Lancet Public Health*, vol. 5, no. 8, Aug. 2020.

[19] M. Salathé, C. L. Althaus, N. Anderegg, D. Antonioli, T. Ballouz, E. Bugnion, S. Čapkun, D. Jackson, S.-I. Kim, J. R. Larus, N. Low, W. Lueks, D. Menges, C. Moullet, M. Payer, J. Riou, T. Stadler, C. Troncoso, E. Vayena, and V. von Wyl, "Early evidence of effectiveness of digital contact tracing for SARS-CoV-2 in Switzerland," *Swiss Medical Weekly*, no. 51, Dec. 2020.

[20] S. Munzert, P. Selb, A. Gohdes, L. F. Stoetzer, and W. Lowe, "Tracking and promoting the usage of a covid-19 contact tracing app," *Nature Human Behaviour*, vol. 5, Feb. 2021.

[21] G. Cencetti, G. Santin, A. Longa, E. Pigani, A. Barrat, C. Cattuto, S. Lehmann, M. Salathé, and B. Lepri, "Digital proximity tracing on empirical contact networks for pandemic control," *Nature Communications*, vol. 12, no. 1, Mar. 2021.

[22] C. Wymant, L. Ferretti, D. Tsallis, M. Charalambides, L. Abeler-Dörner, D. Bonsall, R. Hinch, M. Kendall, L. Milsom, M. Ayres, C. Holmes, M. Briers, and C. Fraser, "The epidemiological impact of the NHS COVID-19 app," *Nature*, vol. 594, no. 7863, Jun. 2021.

[23] N. R. Jones, Z. U. Qureshi, R. J. Temple, J. P. J. Larwood, T. Greenhalgh, and L. Bourouiba, "Two metres or one: what is the evidence for physical distancing in covid-19?" *BMJ*, vol. 370, Aug. 2020.

[24] D. J. Leith and S. Farrell, "Measurement-based evaluation of Google/Apple Exposure Notification API for proximity detection in a light-rail tram," *Plos one*, vol. 15, no. 9, Sept 2020.

[25] P. H. England, "Investigation of novel SARS-CoV-2 variant. Variant of Concern 202012/01 Technical briefing 3. Visited 11/08/2021." [Online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/959360/Variant_of_Concern_VOC_202012_01_Technical_Briefing_3.pdf

[26] World Health Organization and European Centre for Disease Prevention and Control, "Indicator framework for the evaluation of the public health effectiveness of digital proximity tracing solutions," 2021, visited 16/09/2021. [Online]. Available: https://apps.who.int/iris/handle/10665/341818

[27] A. Rimmer, "Sixty seconds on . . . the pingdemic," *BMJ*, vol. 374, July 2021.

[28] M. Vinceti, T. Filippini, K. J. Rothman, S. Di Federico, and N. Orsini, "SARS-CoV-2 infection incidence during the first and second COVID-19 waves in Italy," *Environmental Research*, vol. 197, Jun 2021.

[29] Italian Ministry of Health, "Covid-19 weekly monitoring, report of october 19-25," visited 16/09/2021. [Online]. Available: http://www.salute.gov.it/portale/nuovocoronavirus/dettaglioNotizieNuovoCoronavirus.jsp?lingua=english&menu=notizie&p=dalministero&id=5150

[30] N. Chintalapudi, G. Battineni, and F. Amenta, "Second wave of COVID-19 in Italy: Preliminary estimation of reproduction number and cumulative case projections," *Results in Physics*, vol. 28, Sept 2021.

[31] Q. Bi, Y. Wu, S. Mei, C. Ye, X. Zou, Z. Zhang, X. Liu, L. Wei, S. A. Truelove, T. Zhang, W. Gao, C. Cheng, X. Tang, X. Wu, Y. Wu, B. Sun, S. Huang, Y. Sun, J. Zhang, T. Ma, J. Lessler, and T. Feng, "Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study," *The Lancet Infectious Diseases*, vol. 20, Aug. 2020.

[32] E. Hernández-Orallo, P. Manzoni, C. T. Calafate, and J.-C. Cano, "Evaluating How Smartphone Contact Tracing Technology Can Reduce the Spread of Infectious Diseases: The Case of COVID-19," *IEEE Access*, vol. 8, May 2020.

[33] I. Arisi and E. Mantuano, "Age and gender distribution of covid-19 infected cases in italian population," *https://europepmc.org/article/ppr/ppr211220*, 2020, preprint.

[34] D. Giovanelli, E. Farella, D. Fontanelli, and D. Macii, "Bluetooth-Based Indoor Positioning Through ToF and RSSI Data Fusion," in *IEEE International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Sept. 2018.

[35] D. Menges, H. E. Aschmann, A. Moser, C. L. Althaus, and V. von Wyl, "A Data-Driven Simulation of the Exposure Notification Cascade for Digital Contact Tracing of SARS-CoV-2 in Zurich, Switzerland," *JAMA Network Open*, vol. 4, no. 4, Apr. 2021.

[36] C. Segal, Z. Zhang, B. T. Karras, D. Revere, G. Zane, and J. G. Baseman, "Early Epidemiological Evidence of Public Health Value of WA Notify, a Smartphone-based Exposure Notification Tool: Modeling COVID-19 Cases Averted in Washington State," Jun. 2021, preprint. [Online]. Available: https://www.medrxiv.org/content/10.1101/2021.06.04.21257951v4

[37] D. Klinkenberg, K. Leung, and J. Wallinga, "CoronaMelder – modelstudie naar effectiviteit. Digitaal contactonderzoek in de bestrijding van COVID-19," RIVM Rijksinstituut voor Volksgezondheid en Milieu, Tech. Rep., 2021, visited 22/12/2021, document in Dutch, abstract in English. [Online]. Available: https://rivm.openrepository.com/handle/10029/624996

[38] Z. Ratliff and J. Khoury, "Snarks to the rescue: proof-of-contact in zero knowledge," 2020, preprint. [Online]. Available: https://arxiv.org/abs/2005.12676

[39] K. Hyunjun and K. JeongGil, "Privacy-Preserving Contact Tracing Using Homomorphic Encryption: Poster Abstract," in *ACM Conference on Embedded Networked Sensor Systems*, Nov. 2020.

[40] K. Sinha, P. Majumder, and S. K. Ghosh, "Fully Homomorphic Encryption based Privacy-Preserving Data Acquisition and Computation for Contact Tracing," in *IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)*, Dec. 2020.

[41] Y. An, S. Lee, S. Jung, H. Park, Y. Song, and T. Ko, "Privacy-Oriented Technique for COVID-19 Contact Tracing (PROTECT) Using

Homomorphic Encryption: Design and Development Study"," *Journal of Medical Internet Research*, vol. 23, no. 7, Jul 2021.

[42] J. Zhou, H. Zogan, S. Yang, S. Jameel, G. Xu, and F. Chen, "Detecting Community Depression Dynamics Due to COVID-19 Pandemic in Australia," *IEEE Transactions on Computational Social Systems*, vol. 8, no. 4, 2021.

[43] G. Shrivastava, P. Kumar, R. P. Ojha, P. K. Srivastava, S. Mohan, and G. Srivastava, "Defensive modeling of fake news through online social networks," *IEEE Transactions on Computational Social Systems*, vol. 7, no. 5, 2020.

[44] A. Baker, I. Biazzo, A. Braunstein, G. Catania, L. Dall'Asta, A. Ingrosso, F. Krzakala, F. Mazza, M. Mézard, A. P. Muntoni, M. Refinetti, S. Sarao Mannelli, and L. Zdeborová, "Epidemic mitigation by statistical inference from contact tracing data," *Proceedings of the National Academy of Sciences*, vol. 118, no. 32, Aug. 2021.

[45] N. Fenton, S. McLachlan, P. Lucas, K. Dube, G. Hitman, M. Osman, E. Kyrimi, and M. Neil, "A privacy-preserving Bayesian network model for personalised COVID19 risk assessment and contact tracing," 2020, preprint. [Online]. Available: https://www.medrxiv.org/content/10.1101/2020.07.15.20154286v1

Appendix

A. Data Sources

The data sources are the following ones (URLs in Tab. II):

- $i'(t)$ and $S'(t)$: Immuni dataset on github.com;
- W: SwissCovid app, Health Ministry of Switzerland;
- M: The smartphone ownership survey by the Pew research center (2019);
- R: The Italian statistical Institute (ISTAT);
- $i'(t)$ and $T(t)$: The Italian Health Ministry (IHM), github.com
- R_t: IHM weekly report (the summary is in English, the reports in Italian). When reports provide data for overlapping periods, we used the most recent in Fig. 1.

B. Rescaling i(t)

The Immuni dataset reports the number of daily downloads of the app for the iOS and Android Operating Systems, and the total number of notifications sent on both platforms. The data specification document reports that[10]:

"*The number of notifications is the number of notifications of possible risk exposure generated by the application. The detection is partial since all notifications for iOS devices are detected and only a third of those sent by Android devices have the necessary technology available to detect them safely.*"

This means that if we call $N_{iOS}(t)$ the number of notifications that are served to iOS smartphones (all of them correctly reported) and $N_{Android}(t)$ the number of notifications served to Android smartphones (of which only one third is reported), the total number $N(t)$ of notifications provided by the official statistics is to be interpreted as:

$$N(t) = N_{iOS}(t) + \frac{N_{Android}(t)}{3}$$

As confirmed by a public query to the app developers[11]. We only know $N(t)$ and not its single components, so we rescale $N(t)$ by a factor that takes into account the proportion between the number of downloads for iOS (D_{iOS}) and Android ($D_{Android}$). We then recompute $S'(t)$ as:

$$S'(t) = N(t) * \frac{3 * D_{Android} + D_{iOS}}{D_{Android} + D_{iOS}}$$

C. Modification of the Immuni Configuration

Immuni is open source and the development is carried on on the well-known Github.com platform, so it is possible to observe the changes in some of the app features. The threshold used to detect a close contact is not part of the code, but of a configuration value that is pulled by phones from the Immuni server, so it can not be tracked back in time. On two occasions Github issues were used to discuss the configuration of the sensitivity threshold. The first time[12] it was noted that Immuni uses a very low threshold, trying to capture as many contacts as possible. The second[13] Immuni developers confirmed that this policy was changed on Oct. 28th, using a more conservative threshold.

Leonardo Maccari is an associate professor at the Ca' Foscari University of Venice, Italy. He received a Ph.D in the Faculty of Computer Science Engineering in the University of Florence in 2011 and he was a Marie Curie fellow and an Assistant Professor at the University of Trento (Italy) till 2019. He is an IEEE and ACM member, he co-authored more than 60 publications in refereed conferences, journals and book chapters. He is also among the authors of European and US patents. Among his research interest there are network protocols for large-scale wireless mesh networks, Community Networks, 5G wireless backhaul, interdisciplinary studies and network security.

[10]See https://github.com/immuni-app/immuni-dashboard-data/blob/master/format-regional-trend.md. Visited 11/09/2021.

[11]See https://github.com/immuni-app/immuni-dashboard-data/issues/5. Visited 11/09/2021.

[12]See https://github.com/PersonalDataIO/CoronaRiskScoring/issues/5. Visited 11/09/2021.

[13]See https://github.com/immuni-app/immuni-documentation/issues/112. Visited 11/09/2021.

Data	Link (Visited 11/09/2021)
Immuni data set	https://github.com/immuni-app/immuni-dashboard-data
SwissCovid data set	https://www.experimental.bfs.admin.ch/expstat/en/home/innovative-methods/swisscovid-app-monitoring.html
Pew survey	https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/
ISTAT	http://dati.istat.it/?lang=en
IHM data	https://github.com/pcm-dpc/COVID-19/tree/master/dati-andamento-nazionale
IHM R_t	http://www.salute.gov.it/portale/nuovocoronavirus/archivioNotizieNuovoCoronavirus.jsp?lingua=english\&menu=notizie\&p=dalministero\&tipo=Report\%20settimanale\%20Covid-19\&area=nuovocoronavirus\&notizieEN.page=0

TABLE II
THE URLs OF THE DATA SOURCES